NOT TO BE FURTHER COPIED OR REPRODUCED

(* BATCHES *)

3

Easy Re-expression

		Exh.	Page
Introduction 3-1		1 a	3-5
		2	3-6
3A. Logarithms = logs 3-2			
3B. Quick logs 3-8		3 b	3-10
break tables, defined at 3-9		4	3-11
3C. Quick roots 3-12		5 c	3-14
- quick reciprocals, 3-13 to 3-15		6	3-16
3D. Using roots and reciprocals 3-15			
- reciprocal times, 3-21 to 3-22		7 d	3-17
		8	3-19
3E. Comparisons 3-22		9	3-20
		10	3-23/2-
3F. Counted data 3-34			
3G. Relation among powers and logs (optional) 3-43		11 e	3-25/2
- quick reciprocal roots, 3-47		12	3-28
* * *		13	3-29
3H. How far have we come? 3-50		14	3-31
		15	3-32/3
* * *			
3K. How to think about logs (background) 3-51		16 f	3-36/3
		17	3-38/4
3L. Another look at powers and logs (background) 3-52		18	3-42
* * *		19 g	3-46
		20	3-48
3P. Problems 3-55		21	3-49
		h k	
		22 l	3-53
		23	3-54
		24 p	3-56
		25	3-57
		26	3-58
		27	3-59
		28	3-60
		29	3-61/3

3-10 Quick logs

3-14 Quick roots

3-16 Quick reciprocal

3-49 Quick reciprocal roots

3

Easy Re-expression

We have learned how to collect the values of a batch -- in most cases the natural collection scheme lets us look easily and effectively at the "shape" of the numbers as a whole. There are, however, rather frequent exceptions, as exhibit 6 of chapter 1 and exhibit 5 of chapter 2 illustrate. It is time to learn how to deal with these exceptions.

Only one broad approach seems natural. If the way the numbers were gathered or collected does not make them easy to grasp, we should change them -- preserving as much information as we can use -- into a form where we can grasp them easily. In such cases, we need to re-express the data, something that happens rather often.

Data for us means numbers -- whether given in digits, by tallies, or on a graph. Three broad classes of numbers are important enough to be worth both names and some discussions:

- <u>amounts and counts</u> can never be less than zero and could be arbitrarily large. Heights, powers, areas, distances, numbers of deaths or of people all fall here. The simplest indicator of whether re-expression is likely to help us look at a single batch is the ratio of largest value to smallest value. If this is small, close to 1, re-expression cannot change the appearance of the

batch seriously. If this ratio is quite large, say 100 or more, we are almost sure to need re-expression, just to see what is going on. (More about counts when we reach section 3F.)

- balances, where both positive and negative values occur. Profit-and-loss, actual minus predicted, observed-minus fitted are all natural examples. One can usually think of this sort of data as the difference of two amounts or counts. Re-expression of the balance is rarely helpful, but re-expression of the amounts or counts before the subtraction is sometimes quite helpful.

- counted fractions, percentages, and grades, where the reasonable values are hemmed in on both sides. Re-expression is often very helpful, though the techniques are more special. (For the present we will avoid such cases. We will eventually come to dealing with counted fractions in chapter ?23?. Similar techniques often work with percentages. Grades -- which include A,B,C,...E and -,+,++,+++,++++ as examples -- often respond well to slightly more complex techniques.)

3A. Logarithms = logs

What kinds of re-expression we will most frequently find helpful will depend upon the habits of our fellow men, on which ways men habitually choose to write down the numbers with which we start. Civilized "people" evolved from bears, for example, might have different habits than those evolved

from apes. If so, they would be likely to find other forms of re-expression most useful.

As we have noticed, numbers that are always positive and that are not too closely alike are likely to gain from some form of re-expression. The most likely form is taking logarithms. (The next most likely form is taking square roots, to which will come in turn.)

Logarithms -- or as we shall regularly say, logs -- can be made either simple or mysterious. We shall try to keep them simple.

What do you need to know about logs? In the first instance

- how to find good enough logs fast and easily (see next section).
- that equal differences in logs correspond to equal ratios of raw values. (This means that wherever you find people using products or ratios -- even in such things as price indexes -- you are likely to find that using logs, thus converting products to sums and ratios to differences, is likely to help -- particularly in exploratory data analysis.)

Some readers may like a review of the few simple things they need to know about logs. This they can find in section 3K. There are different kinds of logs, but this makes little difference to us here. Of any two kinds, each must be a constant multiple of the other, so that no more care is

needed than is needed keeping feet and inches straight. We will thus routinely write merely "log" -- meaning the kind of log that makes arithmetic easiest, logs to the base 10, also called common logarithms. (Should we need any other kind, we would make a special note.)

The relation of logs to various powers can also be of interest. (We will come to this in section 3G.)

So much for generalities, let us take up the example of exhibit 6 of chapter 1. If we look back to that display, we see a strong tendency for the leaves to be washed into the corners where the leading digits are 1 or 2. What happens if we look at the logs of the same values? Exhibit 1 shows the result.

Notice that extremes, hinges, and median are given both in logs and in megawatts. It often helps to give both forms when we are displaying re-expressed numbers.

This stem-and-leaf display lets us get a rather good feel for the ultimate sizes of the Bureau of Reclamation's power plants. We see a tendency to spread out to high values that is greater than the tendency to spread to low values. The statement that only plants of 15 megawatts or more are included now begins to be related to something we can see. Clearly the stem-and-leaf in logs is more effective than the stem-and-leaf in raw megawatts.

We can see all this more clearly if we compress the display, as explained in the last chapter, taking one stem for each two starting parts, here each two adjacent first digits. Exhibit 2 shows the result.

exhibit 1 of chapter 3

Ultimate power of hydroelectric plants
of the Bureau of Reclamation
(logs found from powers in megawatts, unit = 0.01)

A) STEM-and-LEAF -- for logs, annotated

		*
1	11	8
2	12	0
3	13	488
7	14	08
11	15	2616
14	16	508
to 16	17	08
✓ 2	18	85
.	19	
to 16	20	80863
11	21	38
9	22	1
8	23	5
7	24	50
5	25	8
4	26	3
	27	
	28	
3	29	5
	30	
2	31	2
	32	
1	33	0

* = does not include plants smaller than 15 megawatts
(118 on this scale)

B) 5-NUMBER SUMMARY -- logs and raw

[34, *HMH*: 1.18, 1.52, 1.86, 2.21, 3.30 log megawatts]

[34, *HMH*: 15, 33, 75, 162, 1974 megawatts]

C) SOURCES

Source: The World Almanac, 1966, page 263.
Their source: U. S. Bureau of Reclamation.
Raw version: Exhibit 6 of chapter 1.

This example also appears in exhibit 5 of chapter 1.

exhibit 2 of chapter 3

The result of squeezing exhibit 1
(see exhibit 1)

A) SQUEEZED STEM-and-LEAF - units 0.01 in logs of megawatts

```
    2      *8:    11:12|8:0
    7             13:14|488:08
 to 14     H2:    15:16|1266:058
  ✓ 4      :M8    17:18|08:58
 to 17             19:20|:03688
   12              21:22|38:1
    9      H5:    23:24|5:05
    6             25:26|8:3
    .             27:28|:
    4             29:30|35:
    2             31:32|2:
    1      *0:    33:34|0:
```

Such an entry as **11:12**|8:0 refers to 118; 120.
Similarly, **13:14**|488:08 refers to 134, 138, 138; 140, 148.

B) 5-NUMBER SUMMARY - logs and raw

[35, *HMH*: 1.18, **1.52**, 1.88, **2.35**, 3.30 log megawatts]

[35, *HMH*: 15, **33**, 75, **225**, 1974 megawatts]

C) REMARK

This example also appears in exhibit 5 of chapter 1.

Compression is not something to be done while the data is being gathered together. In the present example, there is no substitute for exhibit 1 as a first step. (Exhibit 2 cannot be written out until we have all the data gathered.) Once we have it, and notice its unduly stretched out appearance, we can proceed directly to exhibit 2.

3B. Quick logs

We have just seen an example where two-decimal logs were quite precise enough. Indeed, in this example, we could have squeezed through fairly well with one-decimal values. This is but one example of many. Though there are exceptions, in most situations where re-expression on logs really helps, two-decimal logs help as much as many-decimal ones.

Since we can find two-decimal logs easily, as we shall soon see, our sailing instructions are clear:

- plan to use two-decimal logs first always.
- in the rare cases where we need more precision, go back and start again with many-decimal logs.

Overall, we will do much less work this way.

To get few-decimal logs easily, a special kind of tables are very helpful. They make looking up logs just a matter of finding our place and reading the answer. (No interpolation!)

For one-decimal logs, we can use the first two columns of the following little table, whose use is easily learned. (It gives only the fractional part of the log, you have to find the integer part as usual.)

First nonzero digits of x	First decimal of log x	Break
8913 to 1122	.0	8913
1123 to 1412	.1	1122
1413 to 1778	.2	1413
1779 to 2238	.3	1778
2239 to 2818	.4	2239
2819 to 3548	.5	2818
3549 to 4466	.6	3549
4467 to 5623	.7	4466
5624 to 7079	.8	5624
7080 to 8912	.9	7079
		8913

If we use the standard code -- when in doubt go to an even number -- we can compress this table quite a lot further. We have only to use the last two columns, entering between the breaks.

Tables that work like this are often called critical tables. We will try to be more specific and call them

break tables.

Exhibit 3 uses two break tables to make finding two-decimal logs easy.

Let us give an example of how this works in practice. Exhibit 4 gives the ultimate power in megawatts of the hydro-electric stations and power plants of the Corps of Engineers, first as raw values and then as (unsorted) logs. Let us trace a few examples. 14 comes between 1396 and 1429, so its two-decimal log ends in .15. Similarly, 18 comes between 1799 and 1841, so its two-decimal log ends in .26. Note the entry -- 15,26, -- of these logs in the first line of panel B.

exhibit 3 of chapter 3

Break table for two-decimal logs

A) MAIN BREAK TABLE

Break	log	Break	log	Break	log	Break	log	Break	log
9886	.00	1567	.20	2483	.40	3936	.60	6237	.80
1012	.01	1603	.21	2541	.41	4027	.61	6383	.81
1035	.02	1641	.22	2600	.42	4121	.62	6531	.82
1059	.03	1679	.23	2661	.43	4217	.63	6683	.83
1084	.04	1718	.24	2723	.44	4315	.64	6839	.84
1109	.05	1758	.25	2786	.45	4416	.65	6998	.85
1135	.06	1799	.26	2851	.46	4519	.66	7161	.86
1161	.07	1841	.27	2917	.47	4624	.67	7328	.87
1189	.08	1884	.28	2985	.48	4732	.68	7499	.88
1216	.09	1928	.29	3055	.49	4842	.69	7674	.89
1245	.10	1972	.30	3126	.50	4955	.70	7852	.90
1274	.11	2018	.31	3199	.51	5070	.71	8035	.91
1303	.12	2065	.32	3273	.52	5188	.72	8222	.92
1334	.13	2113	.33	3350	.53	5309	.73	8414	.93
1365	.14	2163	.34	3428	.54	5433	.74	8610	.94
1396	.15	2213	.35	3508	.55	5559	.75	8810	.95
1429	.16	2265	.36	3589	.56	5689	.76	9016	.96
1462	.17	2317	.37	3673	.57	5821	.77	9226	.97
1496	.18	2371	.38	3758	.58	5957	.78	9441	.98
1531	.19	2427	.39	3846	.59	6095	.79	9661	.99
1567		2483		3936		6237		9886	

When in doubt use an even answer, thus 1462 gives .16 and 1496 gives .18.

B) SETTING DECIMAL POINTS

1	+0	-1	1
10	+1	-2	0.1
100	+2	-3	0.01
1000	+3	-4	0.001
10,000	+4	-5	0.0001
100,000	+5	-6	0.00001
1,000,000			0.000001

C) EXAMPLES

Number	B		A		log number
log 137.2	2	+	.14	=	2.14
log 0.03694	-2	+	.57	=	-1.43
log 0.896	-1	+	.95	=	-0.05
log 174,321	+5	+	.24	=	5.25

exhibit 4 of chapter 3

Re-expression of the ultimate power in megawatts of Hydroelectric stations and power plants of the Corps of Engineers

A) STEM-and-LEAF

to 2	**1**	48
to 4	**2**	86
to 9	**3**	60040
to 11	**4**	35
to 13	**5**	42
to 17	**6**	0881
to 21	**7**	5606
to 22	**8**	6
to 24	**9**	06
to 38	**1**	see below
to 20	**2**	50,80,04,00,70
to 15	**3**	40,20,30,
to 12	**4**	68,00,
to 10	**5**	18,40,95,
to 7	**6**	00,
	7	
to 6	**8**	10,10,
	9	
to 4	**1**	728,400,743
to 1	**2**	700

B) RE-EXPRESSED in LOGS (unit = 0.01

1	15,26,
	45,42,
	56,48,48,53,48,
	63,65,
	73,72,
	78,83,83,79,
	88,88,85,88,
	93,
	95,98,
2	see below
	40,45,31,30,43,
	53,51,52,
	67,60
	71,73,77
	78,
	91,91,
3	24,15,24,
	43,

1|10,30,12,00,35,00,24,75,18,65,00,00,35,40,00,00,10,30

becomes

2|04,11,05,00,13,00,09,24,07,22,00,00,13,15,00,00,04,11,

C) 5-VALUE SUMMARIES

[62, *HMH*: 14, **68**, 110, **280**, 2700 megawatts]

[62, *HMH*: 1.15, **1.83**, 2.04, **2.45**, 3.43 log megawatts]

D) SOURCE

The World Almanac, 1966, and Book of Facts, page 265.
Their source: Corps of Engineers, U. S. Army.

Since raw values from 10 to 99 -- the listed values go from 14 to 96 -- will all have logs beginning with "1", while those from 100 to 999 will begin with "2", etc, the entries in panel B fall into a stem-and-leaf like pattern, where a whole block of leaves go with a single starting digit. Moving on farther, 518 becomes between 5070 and 5188, so its log ends in .71. We see the appropriate entry in panel B, corresponding to a log of 2.71. And so on.

To complete the process and see what these ultimate powers are like, we have only to sort out the logs into a conventional stem-and-leaf, after which we are likely to want to condense it to two starting parts per stem. (We leave this to the reader.)

3C. Quick roots

We are also likely to want to find quick square-roots. Here we can still do well with a break table, although the situation is not quite as nice, since we can't reuse a square-root table until x has changed by a factor of 100 -- not 10, as for logs -- as we see from the following simple example:

$$\sqrt{2} = 1.42$$
$$\sqrt{20} = 4.47$$
$$\sqrt{200} = 14.2$$

We could get two-figure accuracy in our roots -- we shall drop the "square-" except where it is really needed -- with a table of the same form as exhibit 3, but we would not

be likely to like it when we had it. Going from

$$\sqrt{100} = 10 \quad \text{to} \quad \sqrt{121} = 11$$

and from

$$\sqrt{10000} = 100 \quad \text{to} \quad \sqrt{12100} = 110$$

is a big step. Going from

$$\sqrt{96{,}04} = 9.8 \quad \text{to} \quad \sqrt{98.01} = 9.9$$

and from

$$\sqrt{9604} = 98 \quad \text{to} \quad \sqrt{9801} = 99$$

is a much smaller step -- only about one-tenth the size, however you figure it. These would be the smallest steps in the answers from an exactly two-figure table of square roots.

We can do better by taking smaller steps near 11 and larger steps near 98. Exhibit 5 is a break table that does just this.

This table will be put to use in the next section.

* quick reciprocals *

From time to time, we find re-expressing numbers by reciprocals useful. When we do this, it is often convenient to preserve order, to have the more positive raw value correspond to the more positive reciprocal. This doesn't happen automatically, in fact quite the reverse. Thus 3 is greater than 2, while 1/3 -- the reciprocal of 3 -- is less than 1/2 -- the reciprocal of 2.

exhibit 5 of chapter 3

Break table for $\sqrt{x}$

A) EXAMPLES

Start by dividing number into periods of two digits, decimal point falling between periods. Thus 124.2 is 1 24 2, but 1242 is 12 42. Likewise 0.00654 is 00 65 4 or 65 4.

Number	Periods	from B	from C	Number
124.2	1 24 2	ab.	112	11.2
1242	12 42	ab.	35	35.
.00654	65 4	.0x	80	.080

B) BREAK TABLES SET DECIMAL POINT -- enter between bold figures, leave with light figures

1	a	.x	**1**
1 00	ab.	.0x	**.01**
1 00 00	abc.	.00x	**.00 01**
1 00 00 00	abcd.	.000x	**.00 00 01**
1 00 00 00 00			**.00 00 00 01**

C) MAIN BREAK TABLE -- in and out as in panel B

Break	$\sqrt{x}$	Break	$\sqrt{x}$	Break	$\sqrt{x}$	Break	$\sqrt{x}$	Break	$\sqrt{x}$
98 01	100	2 49 64	160	5 66	240	15 60	40	35 40	60
1 02 01	102	2 62 44	164	5 86	246	16 40	41	37 21	62
1 06 09	104	2 75 56	168	6 20	252	17 22	42	39 69	64
1 10 25	106	2 89 00	172	6 50	258	18 06	43	42 25	66
1 14 49	108	3 02 76	176	6 81	264	18 92	44	44 69	68
1 18 81	110	3 16 84	180	7 13	270	19 80	45	47 61	70
1 23 21	112	3 31 24	184	7 45	276	20 70	46	50 41	72
1 27 69	114	3 45 96	188	7 78	282	21 62	47	53 29	74
1 32 25	116	3 61 00	192	8 12	288	22 56	48	56 25	76
1 36 89	118	3 76 36	196	8 47	294	23 52	49	59 79	78
1 41 61	120	3 92 04	200	8 82	30	24 50	50	62 41	80
1 48 84	124	4 08 04	204	9 30	31	25 50	51	65 61	82
1 58 76	128	4 24 36	208	9 92	32	26 52	52	68 89	84
1 69 00	132	4 41 00	212	10 56	33	27 56	53	72 25	86
1 79 56	136	4 57 96	216	11 22	34	28 62	54	75 69	88
1 90 44	140	4 75 24	220	11 90	35	29 70	55	79 21	90
2 01 64	144	4 92 84	224	12 60	36	30 80	56	82 81	92
2 13 16	148	5 10 76	228	13 32	37	31 92	57	86 49	94
2 25 00	152	5 29 00	232	14 06	38	33 06	58	90 25	96
2 37 16	156	5 47 56	236	14 82	39	34 22	59	94 09	98
2 49 64		5 66 44		15 60		35 40		98 01	

The way out is simple:

use negative reciprocals

These work well, since -1/3 is more positive than -1/2.

Exhibit 6 gives a break table for negative reciprocals. Except for the minus sign, no new feature. To look up -1/x for x = 43.7, we note that 437 comes between 4347 and 4425, so that the nonzero figures of -1/x are -228 (to this accuracy). Fixing the decimal point using panel A gives us -.0228 for -1/x.

3D. Using roots and reciprocals

Examples where stem-and-leaf displays of roots clearly give us the best grasp are harder to find than examples where this happens for logs -- the latter seem to turn up all over.

Exhibit 7 sets out the heights of those volanoes that the National Geographic Society and the 1966 World Almanac found interesting enough to list. Clearly we have more trailing off to large heights than to small ones. Exhibit 8 shows the result of taking logarithms; We now have more trailing toward low logs than we have toward high logs. Clearly we want something in between raw values and logs. Square roots are just the thing.

Exhibit 9 shows the stem-and-leaf for the roots. Now we have a nice symmetric picture, one that fits in well with a median of nearly 6500 feet and extremes near 200 and 19,900. If one wanted to compare interesting volcanoes

exhibit 6 of chapter 3

Critical table for -1/x

A) BREAK TABLES for SETTING DECIMAL POINT

1	.x	a.	1
10	.0x	ab.	.1
100	.00x	abc.	.01
1000	.000x	abcd.	.001
10,000	.0000x	abcde.	.0001
100,000	.00000x	abcdef.	.00001
1,000,000	.000000x	abcdefg.	.000001
10,000,000			.0000001

Examples

Number	A	B	-1/number
124.2	.00x	-80	-.0080
.04739	ab.	-212	-21.2
1242.	.000x	-80	-00080

B) MAIN BREAK TABLE

Break	-1/x	Break	-1/x	Break	-1/x	Break	-1/x	Break	-1/x
990	-100	1639	-60	2469	-40	4115	-240	610	-160
1010	-98	1681	-59	2532	-39	4202	-236	633	-156
1030	-96	1709	-58	2597	-38	4274	-232	649	-152
1053	-94	1739	-57	2667	-37	4347	-228	666	-148
1075	-92	1770	-56	2740	-36	4425	-224	685	-144
1099	-90	1802	-55	2816	-35	4504	-220	704	-140
1124	-88	1835	-54	2899	-34	4587	-216	725	-136
1149	-86	1869	-53	2985	-33	4672	-212	746	-132
1176	-84	1905	-52	3077	-32	4762	-208	769	-128
1205	-82	1942	-51	3175	-31	4854	-204	793	-124
1235	-80	1980	-50	3287	-30	4950	-200	820	-120
1266	-78	2020	-49	3367	-294	505	-196	840	-118
1299	-76	2062	-48	3448	-288	515	-192	855	-116
1333	-74	2105	-47	3509	-282	526	-188	870	-114
1370	-72	2151	-46	3584	-276	538	-184	885	-112
1408	-70	2198	-45	3663	-270	549	-180	901	-110
1449	-68	2247	-44	3745	-264	562	-176	917	-108
1493	-66	2299	-43	3831	-258	575	-172	935	-106
1538	-64	2353	-42	3922	-252	588	-168	952	-104
1587	-62	2410	-41	4016	-246	610	-164	971	-102
1639		2469		4115		633		990	

exhibit 7 of chapter 3

Heights of some of the world's volcanoes

A) STEM-and-LEAF -- unit = 100 feet

8	*2	0	98766562
16		1	97719630
39		2	699877665444222110098 50
57	H8	3	876655412099551426
79		4	999884433192943336110 7
to 102		5	976666665544222100977 31
18	M6	6	8986654410777610 65
to 98		7	98855431100652108073
78		8	653322122937
66	H6	9	377655421000493
51		10	0984433165212
38		11	4963201631
28		12	45421164
20		13	47830
15		14	00
13		15	676
10		16	52
8		17	92
6		18	5
5	*9	19	39730

[218, *HMH*: 0.2, 3.8, 6.6, 9.6, 20 thousand feet]

M6 is the "round to even" version of $5\frac{1}{2}$.

B) SOURCE

The World Almanac, 1966, and Book of Facts, pages 282-283.
Their source: National Geographic Society.

(page 3-18 has been removed)

exhibit 8 of chapter 3

Logarithmic heights of volcanoes
(from exhibits 7 and 3)

A) STEM-and-LEAF -- units 0.01 in log height in hundreds of feet

1	*0	3	0
.		4	
.		5	
.		6	
5		7	8808
6		8	5
8		9	50
10		10	40
11		11	1
16		12	83380
27		13	88844422000
42		14	266533220650989
57	H8	15	876644319944316
to 79		16	999884433192943336 1107
✓ 27		17	7655555544332221007 66219898
to 112	M1	18	3432211133322199887655558865 59596
79	H8	19	0332211111524799888766555797
51		20	0432211032101686554565599 9889
22		21	00344215599
11		22	0215479998
1	*0	23	0

[218, *HMH* = 0.3, 1.6, 1.8, 2.0, 2.3 log hundreds of feet]

exhibit 9 of chapter 3

Square roots of heights of volcanoes
(from exhibits 7 and 5)

A) STEM-and-LEAF -- units 0.1 of roots of heights in hundreds of feet (equivalently, roots of height in feet)

1	*4	1	4
7		2	865525
11		3	0362
27		4	4114099977766555
50		5	14432211043099867599687
to 72	H2	6	2000220666644666684448
✓ 34		7	0000000664444444444222200666228888
to 112	M0	8	2422200022222088866664448648448486
78	H8	9	2222000004246888886664446 6
52		10	0044222202200068664686
29		11	0222000226886488
13		12	44488
8		13	2266
4	*0	14	0000

....

[218, *HMH*: 14, 62, 80, 98, 140, $\sqrt{\text{feet}}$]

with other interesting mountains, or to compare the interesting volcanoes of the northern and southern hemispheres -- in each case as to height -- we would do well to look at square roots of heights.

* reciprocal times *

It often seems natural to measure times for things to happen. How long does it take a rat -- in training to learn a maze -- to run through the maze (and get his reward)? How long does it take individual fish to stop swimming when exposed to a given concentration of poison? How long after the earthquake did this house fall down? A major difficulty with using times is the rat who never finishes the maze, the fish that never stops swimming, the house that never falls down.

Clearly the rat who never finishes the maze is much more like the slow rat who takes 1000 seconds than either is like the very fast rat who runs in four seconds. Indeed, the 1000-second rat is very much like the 800-second rat, much more so than the 204-second rat is like the four-second rat. The visible difficulty with "never" cases is a warning of a concealed difficulty with "very long time" cases.

We ought not to be using times in such cases, rather we should be using the reciprocal of time. For the rats we have mentioned as illustrations, and some others, we get the following sorts of results:

time	reciprocal time	1000/time
"never"	-0.000	0
1000 sec.	-0.001	1
800 sec.	-0.001_2	1.2
204 sec.	-0.049	49
10 sec.	-0.100	100
6 sec.	-0.167	167
5 sec.	-0.200	200
4 sec.	-0.250	250

If we want to continue to talk about slowness, the negative reciprocals are much better than the time. If we are willing to shift to fastness instead, the 1000/time values without the minus signs work very well indeed.

Exhibit 10 shows examples from a rat-running study by Hull. Panel B shows two long times for the first pair of rats and an overall tendency for the second pair of rats to take less time. The pattern scatter is somewhat uncertain. Panel D shows the stem-and-leaf patterns for the reciprocal times. The first pair of rats now shows a roughly symmetric scatter, while the scatter for the second pair is not symmetrical, although only the one (fastest) value at 340 seems to be straying. The general pattern -- and the comparison -- seems clearer for the reciprocal times.

3E. Comparisons

One notable use of stem-and-leaf displays is for comparison. Exhibit 11 compares the land areas of the 83

Exhibit 10 of chapter 3

Times and reciprocal times for rats from start through section 2 of a runway

A) TIMES in SECONDS -- five trials per rat on second day of training

Rat 0	Rat 1	Rat 2	Rat 3
7.6	11.9	10.8	5.6
12.7	18.6	3.9	7.0
26.1	9.3	6.5	8.1
13.7	22.4	2.9	5.7
7.4	12.8	5.9	4.6

B) STEM-and-LEAF for PAIRS of RATS -- unit 1 second

Rats 0 and 1

0	
	879
1	34293
2	2
	6

Rats 2 and 3

0	43
	6667865
1	1
2	

C) RECIPROCAL TIMES for DATA of A -- in 1000/seconds

Rat 0	Rat 1	Rat 2	Rat 3
132	84	92	180
78	54	252	144
38	108	152	124
72	45	340	176
136	78	168	216

(Cont'd)

exhibit 10 (Cont'd)

D) (COMPRESSED) STEM-and-LEAF for PAIRS of RATS -- in 1000/seconds

Rats 0 and 1

3:4	8:5
5:6	4:
7:8	288:4
9:10	:8
11:12	:
13:14	26:

[10*HMH*: 38,54,78,108,136]

Rats 2 and 3

3:4	:
5:6	:
7:8	:
9:10	2:
11:12	:4
13:14	:4
15:16	2:8
17:18	6:0
19:20	:
21:22	6:
23:24	:
25:26	2:
27:28	:
29:30	:
31:32	:
33:34	:0

[10*HMH*: 92,144,173,216,340]

E) NOTES

Source: Clark L. Hull 1934. The rat's speed-of-locomotion gradient in the approach to food. J. of Comparative Psychology 17: 393-422. (Data from table 3 on page 401.)

exhibit 11 of chapter 3

The land areas of the counties of Michigan and Mississippi

A) SIDE-by-SIDE STEM-and-LEAF - unit 10 sq. miles

Michigan, 83 counties		Mississippi, 82 counties	
3	2	3	
3	75	3	8
4	1	4	01212431223011214202
4	8596689	4	597886556569
5	1042043214	5	142010
5	75877677685776776666678666	5	977899958797
6	4011	6	412441
6	865	6	898598
7	120142	7	320341203
7	65	7	86657
8	3212	8	303
8	668	8	8
9	101	9	24
9	6	9	
10	313	10	
10		10	
11	1	11	
11	8	11	
12	00	12	
12		12	
13	2	13	
13		13	
1	58,84,	1	

(Cont'd)

exhibit 11 (Cont'd)

B) BACK-to-BACK STEM-and-LEAF -- sorted, as well

Michigan		Mississippi
2	3	
75		8
1	4	00011111222222223344
9988665		555566678899
4443221100	5	001124
888777777777766666666666655		577778899999
4110	6	112444
865		588899
422110	7	001223334
65		56678
3221	8	033
866		8
110	9	24
6		
331	10	
81	11	
00	12	
2	13	
	14	
	15	
8		
	16	
	17	
4	18	
(83, check)		(82, check)

C) SOURCE

The World Almanac, 1966, pages 369-370.

counties of Mississippi. We see that county areas in Michigan:

- tail out much more toward large areas
- are strikingly concentrated between 545 and 595 square miles (26 out of 83).

Panel B shows another way of displaying the same information. Some will prefer the one, others the other.

Let us look to see what is the effect of going over to logs. Exhibit 12 shows the same comparison for logs. On the whole, the comparison is clearer:

- setting aside the 23 Michigan counties with logs of 1.75 or 1.76, the rest of the Michigan counties appear as a rather triangular pattern -- running from, say, 1.6 (blunt end) to 2.2 (point) -- while the pattern of the Mississippi counties appears rather rectangular -- running from near 1.6 to near 1.9.

Here the gain from re-expression came mainly from improving the picture for the separate stem-and-leaf pictures.

Sometimes, the gain from re-expression comes from matching the stem-and-leaves better to each other. Exhibit 13 shows the "shoreline" of nine states with a medium amount of shoreline according to each of three definitions (all from the U.S. Coast and Geodetic Survey):

- general coastline -- "measurements made with a unit measure of 30 minutes of latitude"
- general tidal shoreline -- "measurements made with a unit measure of three statute miles"

exhibit 12 of chapter 3

Logs of areas of counties of Michigan and Mississippi

A) BACK-to-BACK STEM-and-LEAF -- unit = 0.01 in log

Michigan		Mississippi
41	**15**	
5		8
1	**16**	00011111222222223344
988665		55556667889
443332221100	**17**	0011234
99866666666666655555555555		66666677777999
3211	**18**	111133344
88766555		55566666788899
4332111	**19**	0224
8665		67
110	**20**	
8875		
2	**21**	
70		
	22	

exhibit 13 of chapter 3

Three ways to measure shoreline

A) STATES INCLUDED:

Maine, Massachusetts, New York, New Jersey, Virginia, South Carolina, Georgia, Atlantic Coast of Florida, Oregon. (These nine have general tidal shorelines between 300 and 800 miles in length.)

B) STEM-and-LEAVES - unit 1 mile

General coastline

1	92,27,30,12,87,00,
2	28,96,
3	99,
4	
5	

General tidal shoreline

1	
2	
3	98,12,
4	53,70,
5	67,
6	76,03,18,
7	58,

Detailed tidal shoreline

1	410,
	519,792,850,
2	344
3	035,478,315

C) SOURCE

Information Please Almanac, 1957, page 309.
Their source: U. S. Coast and Geodetic Survey

- detailed tidal shoreline -- "measurements made on largest scale maps and charts available in 1939-40".

On exhibit 13 we can make a low-grade comparison of the first two measures. We can -- and do -- show them to the same scale, though their appearance is rather dissimilar. There is no chance of showing the third, detailed tidal shoreline, measure on any scale at which the general coastline measure would be usefully scratched down.

Exhibit 14 shows the logs of these lengths. We can now display to a single scale and make better comparisons. The spreads of the three stem-and-leaf displays are now at least of the same general size -- from one extreme to the other they are now 0.60, 0.39, and 0.39 rather than 299 mi, 446 mi, and 1968 mi. This display shows how re-expression can -- reasonably frequently -- help by making spreads about the same size.

A different sort of example can be had by turning to what was known about the atomic weights of the chemical elements in 1927. The values in exhibit 15 are the results found by different investigators in the years 1922 to 1925 which were the basis for the German 1927 standard values for the atomic weights of hydrogen and bromine. Clearly, any attempt to compare by stem-and-leaves with equal units fails for the raw values.

When we turn to logs, we find that five-decimal logs are not good enough, so we use six-decimal values.

exhibit 14 of chapter 3

Logs of measures of coastlines

(from exhibit 13)

A) SIDE-by-SIDE STEM-and-LEAF DISPLAYS -- units 0.01 in log

General Coastline		General Shoreline		Detailed Shoreline	
20	05	20		20	
21	01	21		21	
22	78	22		22	
23	6	23		23	
24	7	24	9	24	
25		25		25	
26	0	26	0667	26	
27		27	589	27	
28		28	38	28	
29		29		29	
30		30		30	
31		31		31	58
32		32		32	57
33		33		33	7
34		34		34	68
35		35		35	24
*	2.00	*	2.49	*	3.15
H	2.10	H	2.66	H	3.25
M	2.27	M	2.75	M	3.37
H	2.36	H	2.79	H	3.48
*	2.60	*	2.88	*	3.54
trt	2.25	trt	2.74	dh	3.37
dh	0.26	dh	0.33	dh	0.23

exhibit 15 of chapter 3

Atomic weights of hydrogen and bromine regarded as best available in 1927

A) RAW ATOMIC WEIGHTS

Hydrogen	Bromine
1.00779	79.916
1.00781	79.920
1.00777	79.920
1.00782	79.927
1.00775	79.912
1.00766	79.926
1.00779	79.915
1.00769	79.915
1.00783	79.918
	79.909

B) STEM-and-LEAF for RAW VALUES -- unit = 0.001, values cut

Hydrogen	Bromine
100|777777777	7990|9
	7991|62558
	7992|0067

C) LOG ATOMIC WEIGHTS

Hydrogen	Bromine
.003370	1.902634
378	656
362	656
382	748
353	612
314	688
370	628
327	628
387	645
	596

(Cont'd)

exhibit 15 (Cont'd)

D) COMPARATIVE STEM-and-LEAF -- log values, unit = .000001

Hydrogen		Bromine	
00331	4	190259	6
00332	7	190260	
00333		190261	2
00334		190262	88
00335	3	190263	4
00336	2	190264	5
00337	080	190265	66
00338	27	190266	
		190267	
		190268	8
		190269	
		190270	
		190271	
		190272	
		190273	
		190274	8

E) NOTES

(1) The comparison of panel B shows us very little.

(2) In panel D the bulk of the two batches are about equally spread out, and we see clearly the tendency of some (i.e. one or two) individual values to stray away from most others.

Source: Landolt-Bornstein Tabellen -- (Supplement to the 5th edition, 1927, page ???

Panel D, which compares log atomic weights in units of the sixth decimal, shows that stem-and-leaf pictures with a common unit can go do well in comparing these batches, if applied to the right expression. Moreover, they can reveal the spread and straying characteristics of the original batches.

That log expression works well here should not surprise us. The best measurement accuracy, at a given date, for a given sort of amount is likely to be characterized by fluctuations and errors that are, roughly, a given fraction of that amount. We usually speak of having about so many significant figures. To have about certain number of decimals in the log is usually closer to the truth, because this is closer to a constant fraction.

We find re-expression to make spreads comparable useful in a quite wide variety of situations, involving both crude and refined values. The natural first choices are logs for amounts and roots for counts.

3F. Counted data

Most of our examples have so far involved measured data. We have slighted counted data.

(Some readers feel quite happy re-expressing heights, areas, megawatts, amounts of money, and the like but develop a curious conscience when someone suggests re-expressing the number of times something occurred. All such a "conscience" can do is to make it more likely for them to miss what they could have found.)

When we have to deal with special cases, such as

- rather small counts (especially when zero appears),

- how many out of a given number survived, smiled, were satisfied, etc.,

- counts arranged in special patterns,

there are some refinements we can wisely use in re-expression. (Eventually we will spend some chapters on these techniques -- see chapters 24, 27, 28, etc.)

For the present, though, so long as we deal with reasonably large counts, we may as well plan to re-express batches of counts just as freely, and for all the same reasons as batches of amounts.

Exhibit 16 shows the 1960 population of the 50 states, first in raw counts and then in logs. Clearly it takes re-expression for us to look at these values. We do particularly well with the squeezed stem-and-leaf of panel C.

Exhibit 17 shows a still more extreme example -- the numbers of deaths for selected causes in 1964 (omitting "all other diseases"). What it takes to make a cause selected is a curious mixture of popular interest (acute poliomyelitis), ease of diagnosis (motor vehicle accidents) and importance (arteriosclerotic heart disease). Again we see both the raw value stem-and-leaf and the stem-and-leaf for logs. Panel B provides extra detail about the very smallest values. The condensed stem-and-leaf for logs in panel C gives us our first clear picture. We see:

exhibit 16 of chapter 3

The 1960 populations of the 50 states

A) STEM-and-LEAF -- raw values in tens of thousands

		1	
2	*3	2	39
4		3	93
5		4	5
.		5	
11		6	377138
.		7	
13	H9	8	69
15		9	75
to 21		1	30,79,74,41,77,86,
✓ 7	M46	2	54,76,18,18,33,38,85,
to 22		3	27,94,04,26,10,41,57,97,95
13	H32	4	95,66,32,56,
9		5	15,
8		6	07,
7		7	82,
		8	
6		9	71,58,
4	*678	1	572,008,678,132,
		2	

B) STEM- and-LEAF -- logs of same, unit = 0.01

to 1	*6	13	6
to 2		14	6
to 4		15	39
to 5		16	5
to 6		17	9
to 11		18	00333
to 15	H5	19	3589
		20	
to 17		21	15
to 21		22	4457
to 25	M9	23	4478
to 25		24	04589
to 20		25	1135
to 16	H4	26	0004679
to 9		27	18
to 7		28	9
to 6		29	89
to 4		30	05
		31	
to 2	*2	32	02

(Cont'd)

exhibit **16** (Cont'd)

C) STEM-and-LEAF -- as in B, compressed

2	*6:	**13:14**	6:6
5		**15:16**	39:5
11		**17:18**	9:00333
15	H5:	**19:20**	3589:
to 21		**21:22**	15:4457
✓ 8	M9:	**23:24**	4478:04589
to 20	H:4	**25:26**	1135:0004679
9		**27:28**	18:9
6		**29:30**	89:5
2	:*2	**31:32**	:02

D) 5-VALUE SUMMARIES

[50, *HMH*: 0.23, **0.89**, 2.5, **4.3**, 17 millions)

[50, *HMH*: 1.36, **1.95**, 2.39, **2.64**, 3.22 "logs of tens of thousands"]

[50, *HMH*: 5.35, **5.95**, 6-39, **6.64**, 7.22 "logs of people"]

E) SOURCE

The World Almanac, 1966, page 325
Their source: U. S. Bureau of the Census.

Previous appearance: exhibit 5 of chapter 3 .

exhibit 17 of chapter 3

Deaths for selected causes, 1964

(Total 1,798,051, less "all other diseases 54,600")

A) STEM-and-LEAF -- raw values in hundreds

9	*0	0	731018042 (zeroes are 42 and 17)
13		1	7831
15		2	65
17	H6	3	58
21		4	9464
23		5	94
24		6	5
25		7	6
26		8	2
to 29		9	998
✓ 7	M10	1	35,35,47,59,22,11,10,
to 23	H77	2	62,77,57,34,32,03,52,53,06,
14		3	28,23,72,07,
10		4	92,00,69,
7		5	32,74,78,69,
.		6	
.		7	
.		8	
3		9	32,
2		1	982,
.		2	
.		3	
.		4	
1	*454	5	454,

(Cont'd)

exhibit 17 (Cont'd)

B) DETAILS of SMALL VALUES -- in single deaths

Raw	log	Cause
0		
1 7	1.23	Polio
2		
3		
4 2	1.62	Diptheria
5		
6		
7		
8		
9 53	1.98,1.97,	Scarlet fever, whooping cough
1		
2 83,47,	2.45,2.39,	Dysentery, abortions
3		
4 21,	2.62	Measles
5		
6 57,	2.82	TB (other forms)
7 50,	2.88	Meningococcal infections
8		
9		

(Cont'd)

exhibit 17 (Cont'd)

C) STEM-and-LEAVES -- logs of hundreds, unit = 0.01

loose

-8	
-7	7
-6	
-5	
-4	
-3	8
-2	
-1	
-0	32
+0	
1	
2	
3	9
4	5
5	
6	2
7	
8	28
9	
10	4
11	1
12	36
13	
14	01
15	48
16	4469
17	37
18	18
19	1999
20	449
21	337
22	0
23	1177
24	001249
25	127
26	079
27	3666
28	
29	7
30	
31	
32	
33	0
34	
35	
36	
37	4

compressed

-8:-7	:7
-6:-5	:
-4:-3	:8
-2:-1	
-0:+0	32:
1:2	
3:4	9:5
5:6	:2
7:8	:28
9:10	:4
11:12	1:36
13:14	:01
15:16	48:4469
17:18	37:18
19:20	1999:449
21:22	337:0
23:24	1177:001249
25:26	127:079
27:28	3666:
29:30	7:
31:32	
33:34	0:
35:36	
37:38	4:

D) SOURCE

The World Almanac, 1966, page 299.
Their source: National Center for Health Statistics.

- a pronounced tendency to trail off toward low values.

- a weak tendency to trail off toward high values.

- a tendency toward asymmetry, even with these extreme trailing values eliminated, the lower values being stretched out further.

Even if we disregard the extreme trailing values, the use of square roots deserves a trial. This is done in exhibit 18. We find that, even in the compressed form, that we still do not get too clear a picture of all these values as a batch. The tendency of the three highest values -- for "malignant neoplasm of digestive organs and peritoneum", "vascular lesions affecting central nervous system", "arteriosclerotic heart disease" -- to appear disconnected from the main body of causes is now more marked.

We are thus led to think that this particular collection of causes of death includes:

- a number of unusually small numbers, associated with infectious diseases.

- a general mass of medium-sized numbers.

- three unusually large values, which seem rather like "catchalls".

By re-expressing our counts we have learned something about even this unpromising set of data.

exhibit 18 of chapter 3

Deaths for selected causes, 1964, again

(from exhibit 17)

A) STEM-and-LEAF -- in square roots of numbers of deaths

loose

0	46	
1	0067	10
2	06	14
3	36	23
4	12	to 29
5	01	✓ 5
6	02668	to 25
7	046	23
8	08	19
9	0999	13
10	44	10
11	066	8
12	08	6
13		.
14	44	.
15	22	3
16	00008	2
17	6	.
18	00	.
19	2	1
20	0	
21	6	
22	0	
23	2	
24	000	
2		
3	1q,	
4	5q,	
5		
6		
7	4q,	

compressed

0:1	46:0067
2:3	08:36
4:5	12:01
6:7	02668:046
8:9	08:0999
10:11	44:066
12:13	08:
14:15	44:22
16:17	00008:6
18:19	00:2
20:21	0:6
22:23	0:2
24:25	000:
26:27	
28:29	
3	1q,
4	5q,
5	
6	
7	4q,

B) NOTES

In 1q, 5q, and 4q, etc., the "q" represents an additional digit not provided by the table of exhibit 17. (We could get these from the source for that exhibit.)

3G. Relation among powers and logs (optional)

We have said that logs and square roots are more likely to be useful ways of re-expression than any others. We have seen -- by way of the heights of interesting volcanoes -- that the square root does seem to fit in halfway between raw values and logs. This is enough to take us through most situations.

As preparation for the future, though, we are going to look at a wider variety of powers, and to learn in more detail how logs fit in among the powers.

We are now trying to think -- and learn -- about re-expression in general, not just re-expression for making better stem-and-leaf displays. Thus we should be fairly general, which is likely to mean saying some things that may seem trivial.

Changing the numbers in a batch -- or in some more complicated data structure -- by multiplying or dividing all of them by the same constant usually does not affect the ease with which they can be looked at -- either directly or after most standard arithmetic analyses. To make such a change is to do no more than to change feet to inches or meters, or to change kilowatts to megawatts. We usually regard this sort of change as quite trivial -- so far as easing our analysis goes or making it more difficult. (We do need to keep careful track of what we have done, so that we can find our way back when the analysis is completed.)

We noticed above that the sort of re-expression we are discussing is mainly used for numbers that are all of one sign. If all the numbers we have -- or might reasonably have had -- are of one sign, there is likely to be a special significance for zero.

There are many purposes for which adding or subtracting the same constant to each number in a batch -- or in some other data structure -- is just as trivial as multiplying or dividing by a common constant. The most important exception is this:

- if there is a natural zero for the initial numbers, then, once we have added or subtracted a constant, shifting the zero, we are unlikely to take a power, a root, or a log of the modified numbers -- unless we are dealing with small counts, where shifting "none counted" to a small positive value is often followed by taking a root or a log.

Surely the simplest important change in how something is expressed in numbers is to replace the raw numbers by the same simple power of each of them. It is not usual for this power to be a square or a cube. Because of how people choose to make instruments and write numbers down -- as we have noticed -- we much more frequently change them in the opposite direction, as by taking:

-square roots (1/2 powers).

- reciprocals (-1 powers).

- reciprocals of square roots (-1/2 powers).

(For all these but one we have already seen critical tables.)

A look at these powers

$$-1, \ -1/2, \ +1/2, \ +1$$

where the last entry refers to the raw data itself, reveals an obvious gap. What about the zero power?

We were almost all once taught that anything raised to the zero power was 1. Our teachers were not wrong. But that does not mean that there can be nothing to fill the gap.

It turns out that the role of a zero power is for these purposes neatly filled by the logarithm.

Exhibit 19 shows how $y = (\text{const})\cdot(\log x)$ fits neatly in among the simple powers. Notice that negative powers need negative signs to make y increase as x increases. (For a different sort of picture see section 3L.)

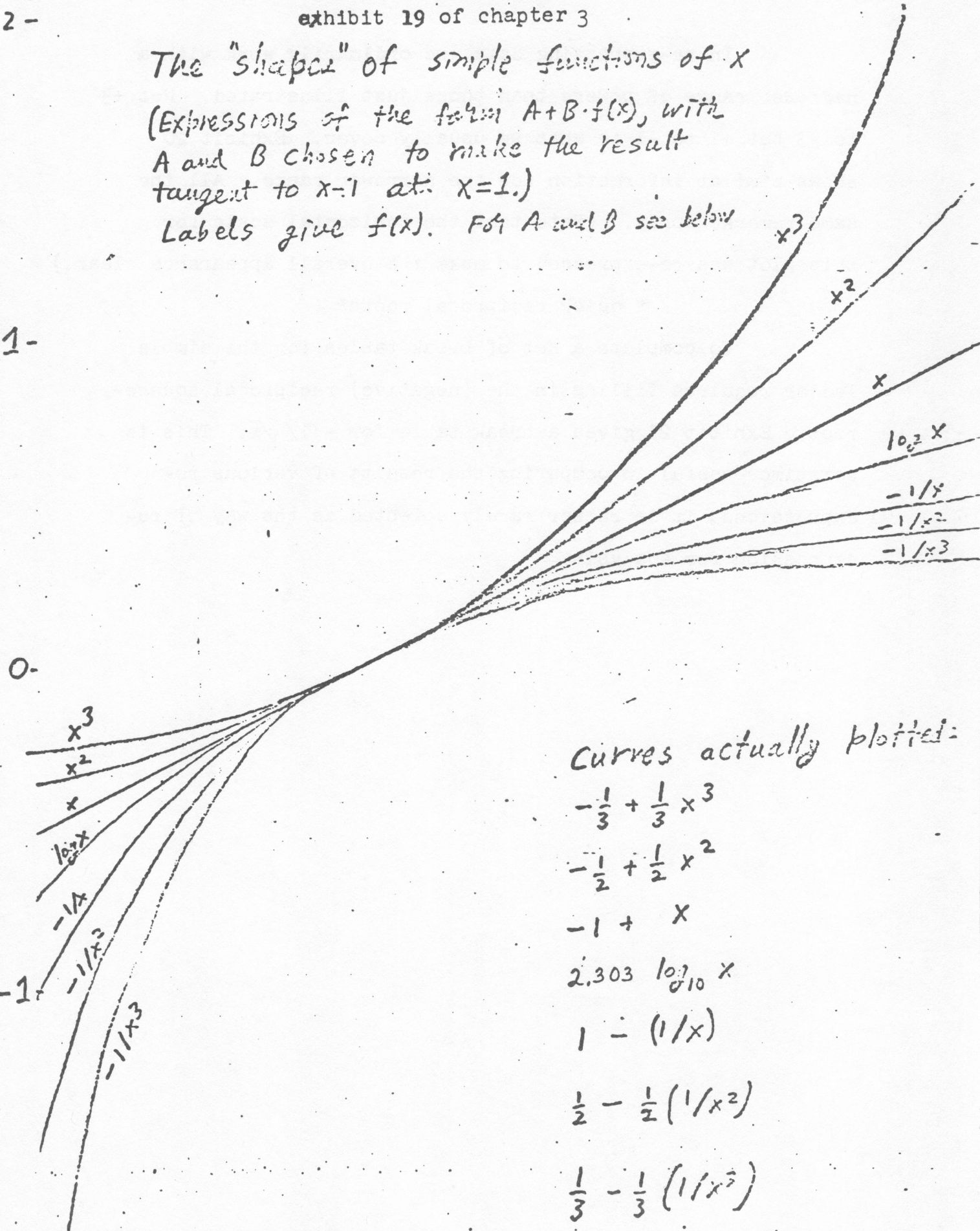
A central point is that moving up the scale -- whether from x to x^2 to x^3 or from $-1/x^2$ to $-1/x$ to log x -- corresponds to emphasizing differences among larger x's in comparison with differences among smaller x's. To do the opposite, emphasizing more strongly differences among smaller x, we need to move down the scale -- as from x to log x to $-1/x$.

exhibit 19 of chapter 3

The "shapes" of simple functions of x

(Expressions of the form $A+B\cdot f(x)$, with A and B chosen to make the result tangent to $x-1$ at $x=1$.)

Labels give $f(x)$. For A and B see below



Curves actually plotted:

$-\frac{1}{3}+\frac{1}{3}x^3$

$-\frac{1}{2}+\frac{1}{2}x^2$

$-1+x$

$2.303\ \log_{10} x$

$1-(1/x)$

$\frac{1}{2}-\frac{1}{2}(1/x^2)$

$\frac{1}{3}-\frac{1}{3}(1/x^3)$

In re-expressing data, we ordinarily work with a narrower range of powers than those just illustrated. Not +3 to -3 but +1 to -1 is what we usually cover. Exhibit 20 shows similar information for the narrower range. All the same remarks apply. (Note that the horizontal scale for this plot was re-expressed to make the overall appearance clear.)

* quick reciprocal roots*

To complete a set of break tables for the simple ladder requires filling in the (negative) reciprocal square-root. Exhibit 21 gives a break table for $-1/\sqrt{x}$. This is sometimes useful in comparing the results of various re-expressions; it is rather rarely selected as the way of re-expression for analysis.

Logs and the usual powers

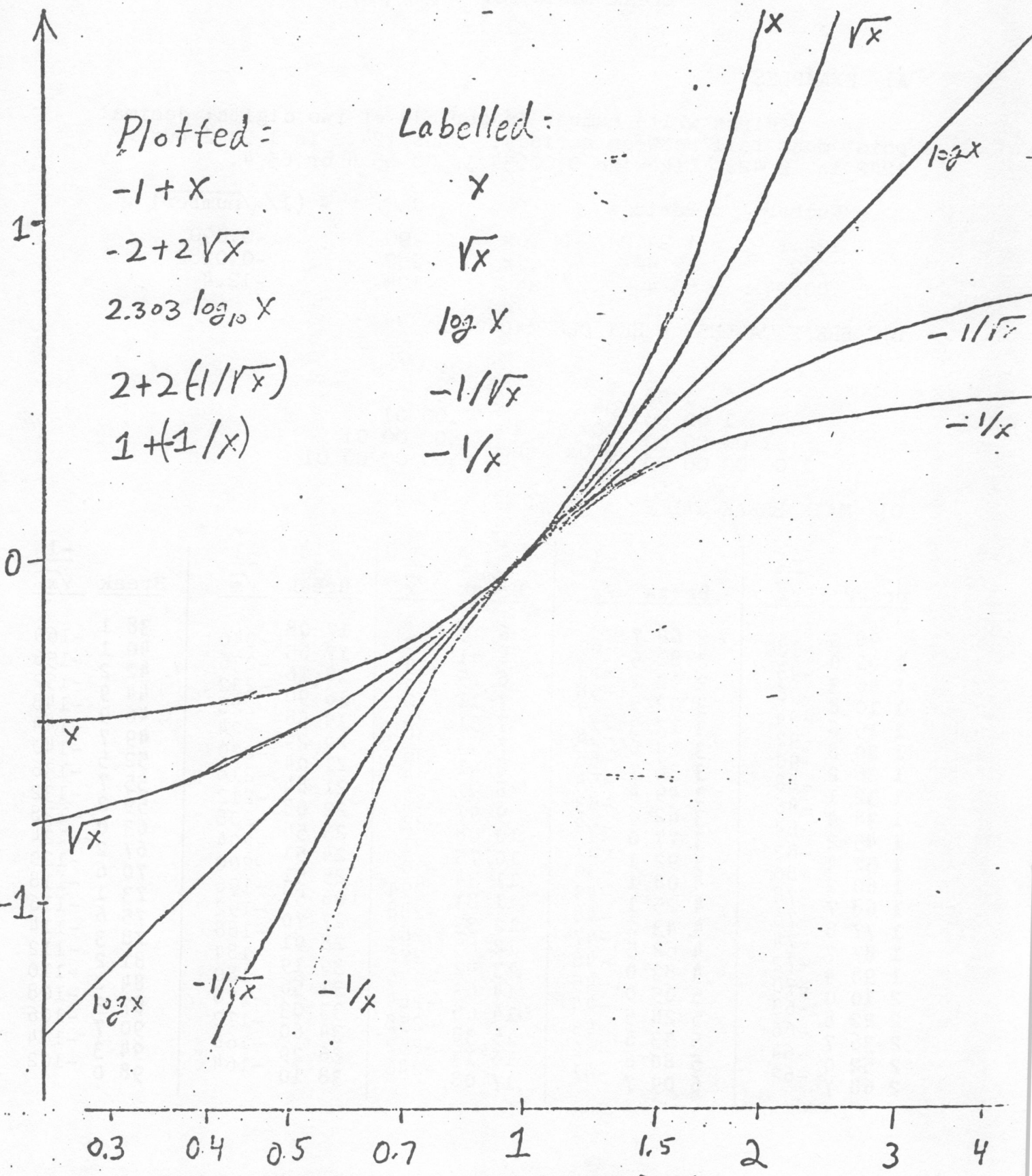


exhibit 21 of chapter 3

Break table for - $(1/\sqrt{x})$

A) EXAMPLES

First write number in periods of two digits; decimal point must fall between periods. Thus 124.2 is 1 24 2 but 1242 is 12 42. Likewise 0.00654 is 00 65 4 or 65 4.

Number	Periods	B	C	$-(1/\sqrt{\text{number}})$
124.2	1 24 2	.0x	-90	-0.090
1242.	12 42	.0x	-282	-0.0282
.00654	65 4	ab.	-124	-12.4

B) BREAK TABLES to SET DECIMAL POINT

1	.x	a.	1.
1 00	.0x	ab.	.01
1 00 00	.00x	abc.	.00 01
1 00 00 00	.000x	abcd.	.00 00 01
1 00 00 00 00			.00 00 00 01

C) MAIN BREAK TABLE

Break	$\frac{-1}{\sqrt{x}}$	Break	$\frac{-1}{\sqrt{x}}$	Break	$\frac{-1}{\sqrt{x}}$	Break	$\frac{-1}{\sqrt{x}}$	Break	$\frac{-1}{\sqrt{x}}$
98 0	-100	2 68 7	-60	6 10	-40	17 08	-240	38 1	-160
1 02 0	-98	2 82 5	-59	6 41	-39	17 65	-236	40 1	-156
1 06 3	-96	2 92 2	-58	6 75	-38	18 26	-232	42 2	-152
1 10 8	-94	3 02 5	-57	7 11	-37	18 90	-228	44 9	-148
1 15 6	-92	3 13 3	-56	7 51	-36	19 58	-224	46 9	-144
1 20 8	-90	3 24 6	-55	7 93	-35	20 29	-220	49 7	-140
1 26 2	-88	3 36 7	-54	8 41	-34	21 04	-216	52 5	-136
1 32 1	-86	3 49 4	-53	8 91	-33	21 84	-212	55 7	-132
1 38 4	-84	3 62 6	-52	9 47	-32	22 68	-208	59 2	-128
1 45 2	-82	3 77 0	-51	10 08	-31	23 56	-204	63 0	-124
1 52 4	-80	3 92 1	-50	10 75	-30	24 51	-200	67 2	-120
1 60 2	-78	4 08 1	-49	11 34	-294	25 51	-196	70 6	-118
1 68 7	-76	4 25 1	-48	11 81	-288	26 57	-192	73 1	-116
1 77 8	-74	4 43 2	-47	12 33	-282	27 70	-188	75 6	-114
1 87 7	-72	4 62 5	-46	12 85	-276	28 91	-184	78 3	-112
1 98 4	-70	4 83 0	-45	13 42	-270	30 19	-180	81 2	-110
2 10 0	-68	5 05 0	-44	14 63	-264	31 56	-176	84 2	-108
2 22 8	-66	5 28 5	-43	14 68	-258	33 03	-172	87 3	-106
2 36 7	-64	5 53 6	-42	15 38	-252	34 60	-168	90 7	-104
2 52 0	-62	5 80 6	-41	16 13	-246	36 29	-164	94 3	-102
2 68 7		6 09 7		17 08		38 10		98 0	

*****

3H. How far have we come?

Data can be broadly divided into:

- amounts and counts. These may need re-expression by powers, roots, or logs just to look at all clearly at single batches when largest/smallest is not small. They may need re-expression to conveniently compare one batch with another in any case.

- balances. Here re-expression is less frequent, and must usually be done before the balances are formed, not afterward.

- counted fractions, percentages and grades. Here re-expression is often essential, but detailed methods for doing it are going to have to wait until we have gone through many chapters.

Break tables make re-expression in the commonest ways:

- first, logarithms.
- second, square roots.
- third, (negative) reciprocals.
- and even, (negative) reciprocal square roots

both easy and quick.

Counted data, unless we are dealing with quite small counts, are handled in the same way as amounts.

Whenever it matters, we do well to treat logs as playing the role of the zeroth power, so far as re-expression goes.

*****

3K. <u>How to think about logs</u> (background)

We want to use logarithms -- "logs" -- freely as a tool. The points that follow make such use easier and are recalled for reference.

(1) Logs can be looked up in tables.

(2) Logs come "to" different bases; however logarithms to different bases only differ by multiplicative constants. As a result, in data analysis, any base is almost (usually exactly) as good as any other.

(3) We shall use "$\log_{10}$" for logs to the base 10 when we need to be sure (and would use "$\log_e$" or "ln" for logs to the base e).

(4) The log of 1 is 0.

(5) The log of zero is not defined.

(6) The log of a negative number can be defined, but is not a real number. (We shall not try to use logs of negative numbers.)

(7) The log of a product is the sum of the logs of the factors, thus

$$\log uvw = \log u + \log v + \log w$$

(8) The log of a ratio is the difference of the logs of numerator and denominator, thus:

$$\log(s/t) = \log s - \log t$$

(9) The log to base 10 of 10 is 1, (that of 100 is 2, of 1000 is 3, of 0.1 is -1, of 0.01 is -2, of 0.001 is -3, etc.).

3L. Another look at powers and logs (background)

In section 3G we plotted against x the results of re-expressing x (see exhibits 19 and 20). Many find these plots the most helpful way of coming to terms with the common re-expressions.

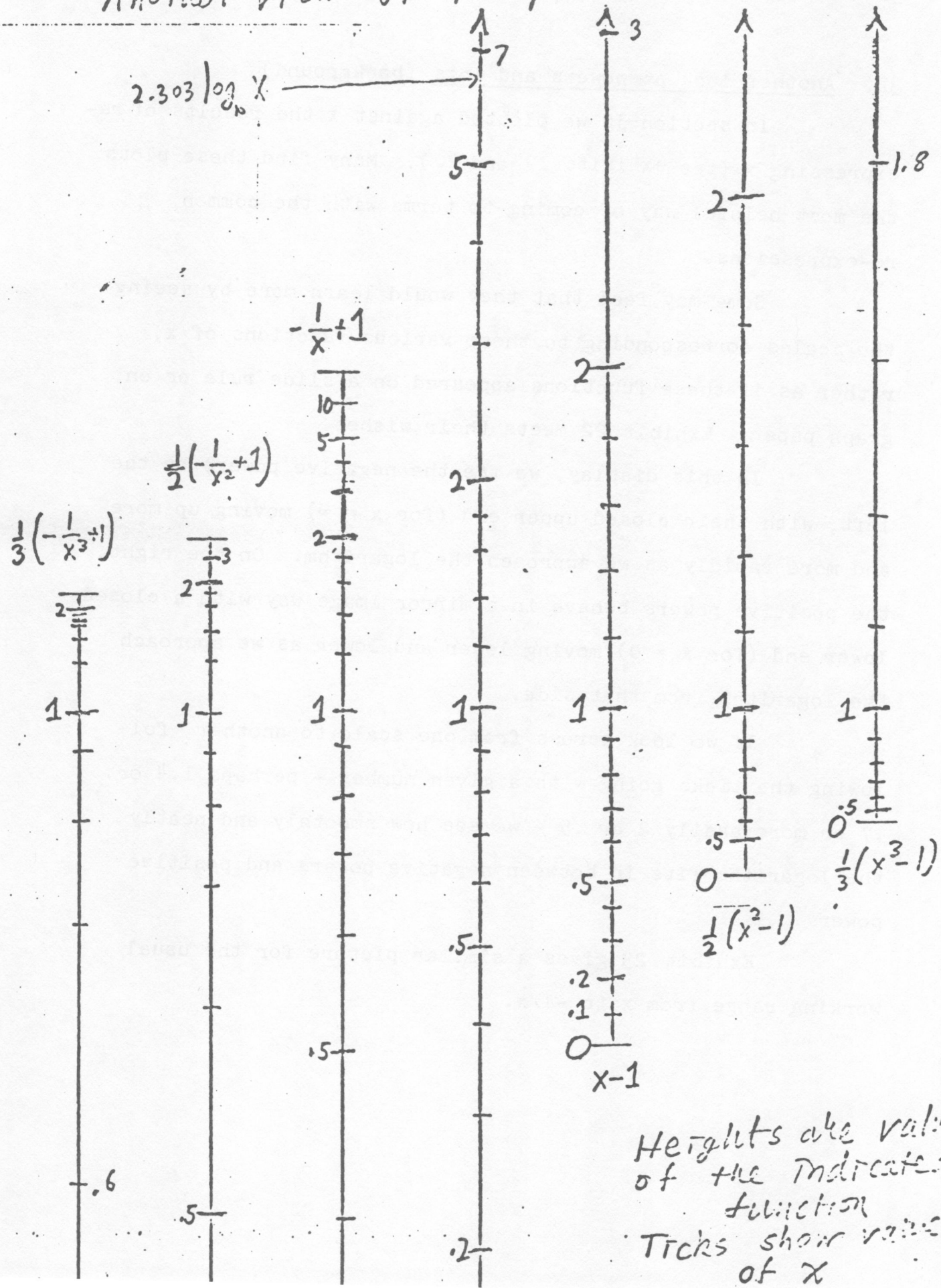
Some may feel that they would learn more by seeing the scales corresponding to these various functions of x, rather as if these functions appeared on a slide rule or on graph paper. Exhibit 22 meets their wishes.

In this display, we see the negative powers on the left, with their closed upper end (for $x = \infty$) moving up more and more rapidly as we approach the logarithm. On the right, the positive powers behave in a mirror image way with a closed lower end (for $x = 0$) moving lower and lower as we approach the logarithm from that side.

If we look across from one scale to another, following the ticks going with a given number - perhaps 1.4 or .7 or more easily 2 or .5 - we see how smoothly and neatly the logarithm fits in between negative powers and positive powers.

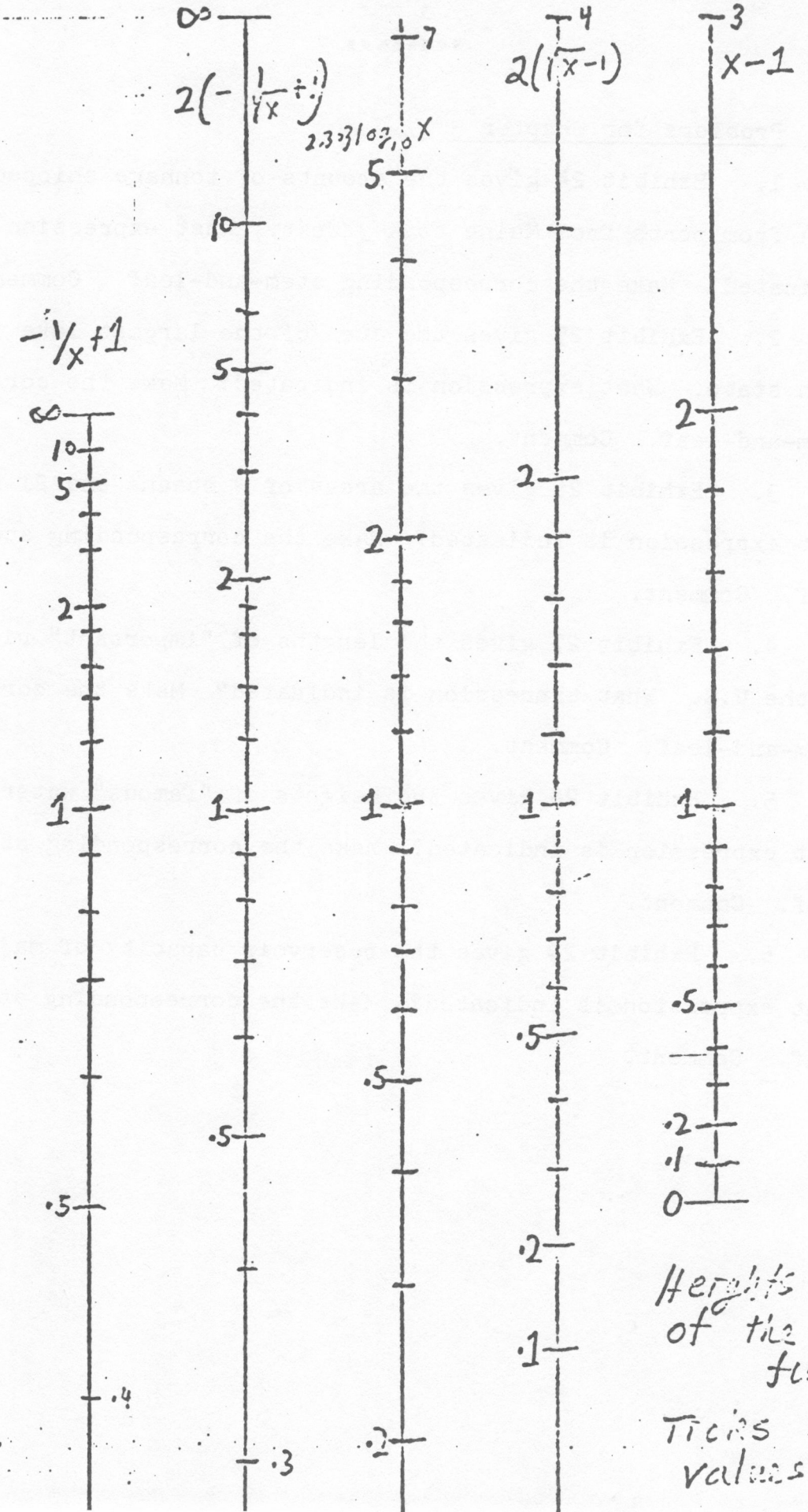
Exhibit 23 gives a similar picture for the usual working range from x to $-1/x$.

Another view of the formulas and of logs



Heights are values of the indicated function
Ticks show values of x

The second view of the usual pairwise and



****

3P. Problems for Chapter 3

1. Exhibit 24 gives the amounts of tonnage shipped in 1964 from ports from Maine to Key West. What expression is indicated? Make the corresponding stem-and-leaf. Comment.

2. Exhibit 25 gives the area of the largest lake in each state. What expression is indicated? Make the corresponding stem-and-leaf. Comment.

3. Exhibit 26 gives the areas of 4 oceans and 21 seas. What expression is indicated? Make the corresponding stem-and-leaf. Comment.

4. Exhibit 27 gives the lengths of "important" rivers of the U.S. What expression is indicated? Make the corresponding stem-and-leaf. Comment.

5. Exhibit 28 gives the heights of "famous" waterfalls. What expression is indicated? Make the corresponding stem-and-leaf. Comment.

6. Exhibit 29 gives the reservoir capacity of major dams. What expression is indicated? Make the corresponding stem-and-leaf. Comment.

exhibit 24 of chapter 3

Tonnage shipped from ports from Maine to Key West

in 1964

(to nearest 1000 tons)

A) The STEM-and-LEAF

```
   8 | 49
   9 | 6
  1  | 63,65,68,71,78.
  2  | 86,
  3  | 18,63,64,71,87,
  4  | 33,58
  5  | 66,
  6  | 24,
  7  | .
  8  | 14,70,
  9  | 09,27,47,60,
 1   |  127,207,249,358,460,771,826,
 2   |  127,223,349,397,699,
 3   |  162,717,
 4   |  158,643,992,
 5   |  107,
 6   |  .
 7   |  161,
 8   |  192,937,
 9   |  232,
1    |   2206,3206,5206,7622,7841,8830,
2    |   0011,
3    |   .
4    |   3919,7042,8220,
0    |    .
1    |    49151,
```

B) The SOURCE

<u>The World Almanac, 1966</u>, Page 766.

exhibit 25 of chapter 3

Largest lakes in each U.S. state

A) The AREAS in SQUARE MILES

```
     0|01444567789
     1|15
     2|059
     3|118
     4|7
     5|1589
     6|03
     7|268
     8|0
     9|4
    1|05,08,11,11,17,23,33,36,37,40,49,57,60,88,92,
    2|15,47,52,
    3|60,82,
    4|30,51,
    5|
    6|09,30,
    7|00,
    8|
    9|
   1|033,500,
  0|9910,
  1|.
  2|2400,
  3|1800,
```

B) NOTES

A lake that is the largest lake partly within each of two states is not doubledcounted.

0 = 0.31 in Connecticut
630 = Lake Pontchartrain
609 = Garrison reservoir
700 = Lake Okeechokee
1633 = Iliamma (Alaska)
1500 = Great Salt Lake
9910 = Lake Erie
22400 = Lake Michigan
31800 = Lake Superior

C) The SOURCE

The World Almanac 1966. Page 285.

exhibit 26 of chapter 3

Areas of oceans and seas

A) The AREAS in 1000's of SQUARE MILES

```
      1|
      2|9
      3|
      4|
      5|
      6|39
      7|
      8|
      9|22
-------------------------
     1|  |63,69,78,80,
     2|  |22,
     3|  |08,89,
     4|  |76,82,
     5|  |90,96,
     6|  |.
     7|  |.
     8|  |76,
     9|  |67,
-------------------------
    1|  |  |063,146,998
    2|  |  |.
-------------------------
   0|  |  |  |5440
   1|  |  |  |.
   2|  |  |  |8356
   3|  |  |  |1839
   4|  |  |  |.
   5|  |  |  |.
   6|  |  |  |3802
```

B) NOTES

4 oceans and 21 seas (including "Malayan seas").

C) The SOURCE

<u>The World Almanac, 1966</u>. Page 275.

exhibit 27 of chapter 3

Lengths of "important" rivers of the US

A) LENGTHS in MILES

```
  1|37,44,40,40,48,10,48,28,10,00,43,31,00,12,
  1|71,78,83,75,50,69,64,60,85,64,90,
  2|02,30,10,17,33,10,37,15,46,
  2|80,90,86,80,50,65,60,90,73,91,81,59,80,60,55,50,50,87,76,[illegible],
  3|15,25,30,36,29,27,06,40,00,32,38,10,00,14,01,
  3|92,50,60,52,50,50,75,80,80,82,50,60,60,
  4|07,09,20,24,25,31,35,44,45,50,60,75,
  5|20,00,05,38,40,45,50,50,60,
  6|00,05,18,25,30,52,87,90,
  7|10,30,35,60,
  8|40,60,62,70,90,
  9|06,81,
 ---------------------------------------
 1|||018,038,171,214,306,360,450,450,885,
 2|||315,348,533,
 3|||710,
```

B) The SOURCE

<u>The World Almanac, 1966</u>. Pages 266-267

example 28 of chapter 3

Heights of "famous" waterfalls

A) HEIGHTS in FEET

```
   0|40,54,65,66,68,70,70,75,90,96,98,
   1|01,09,15,20,25,25,30,30,30,32,40,44,50,51,65,68,86,93,95,98,
   2|00,07,07,13,14,18,20,30,40,45,51,51,56,66,70,75,86,
   3|00,00,08,11,15,17,20,30,30,35,44,45,55,60,70,70,94,94,
   4|00,00,06,27,50,59,59,70,
   5|00,05,08,18,25,40,42,90,94,97,
   6|00,20,26,30,40,50,56,
   7|00,26,41,
   8|20,30,48,80,89,90,
   9|74,84,84,
  1|| 000,100,170,218,250,312,312,325,350,385,430,
  1|| 535,600,612,640,650,696,904,
  2|| 000,425,600,648,
  3|| 110,212,
```

B) NOTE

Both highest fall and total fall may appear for one waterfall.

C) The SOURCE

<u>The World Almanac, 1966</u>. Page 286.

exhibit 29 of chapter 3

Reservoir capacity of major world dams

A) RESERVOIR CAPACITY in 1000's of ACRE-FEET

```
  0 | 8
  1 |
  2 |
  3 | 2
  4 | 1
  5 | 3
  6 | 1
  7 | 00
  8 | 11
  9 |
  0 | 14,37,39,46,48,52,61,86,95,
  1 |
  2 | 19,65,
  3 | 24,
  4 | 87,
  5 | 49,
  6 | 00,02,
  7 | 46,56,
  8 | .
  9 | 30,
  1 | 261,325,375,405,586,709,
  2 | 000,030,030,092,095,106,367,448,717,
  3 | 024,453,468,484,648,789,
  4 | 413,493,500,500,
  5 | .
  6 | 100,550,600,
  7 | 055,060,
  8 | 000,000,512,
  9 | 171,402,730,890,
  1 | 0945,2970,4755,7715,9000,9400,
  2 | 3600,4500,4800,7000,7160,
  3 | 1618,2471,
  4 | 7020,
  5 |
  6 | 2000,
  7 |
  8 |
  9 |
  1 | 15000,27281,45115
```

(cont.)

exhibit 29 (cont'd.)

B) NOTES

8 = ?
32 = Curnera
41 = Zeuzler
115000 = Manicoagan #5
127281 = Saad-El-Aiti (High Aswan)
145115 = Bratsk

C) The SOURCE

The World Almanac, 1966. Page 260.